

Pharmacometrics Models with Hidden Markovian Dynamics

Marc Lavielle

Received: date / Accepted: date

Abstract The aim of this paper is to provide an overview of pharmacometric models that involve some latent process with Markovian dynamics. Such models include hidden Markov models which may be useful for describing the dynamics of a disease state that jumps from one state to another at discrete times. On the contrary, diffusion models are continuous-time and continuous-state Markov models that are relevant for modelling non observed phenomena that fluctuate continuously and randomly over time. We show that an extension of these models to mixed effects models is straightforward in a population context. We then show how the Forward-Backward algorithm used for inference in hidden Markov models and the extended Kalman filter used for inference in diffusion models can be combined with standard inference algorithms in mixed effects models for estimating the parameters of the model. The use of these models is illustrated with two applications: a hidden Markov model for describing the epileptic activity of a large number of patients and a stochastic differential equation based model for describing the pharmacokinetics of theophyllin.

Keywords Hidden Markov Model · Diffusion model · Stochastic Differential Equation · Mixed effects model · Epilepsy · Pharmacokinetics

1 Introduction

Markov models are stochastic models used to model changing systems that moves sequentially from one state to another and where each transition to a next state is regarded as random and depends only on its most recent values. In particular, a Markov chain with memory 1 assumes that only the current state is necessary to predict the future state [28].

Markov models are therefore well-suited for diseases for which multiple states may exist. In particular, Markov models may describe the transitions between different states in disorders such as depression, migraine or epilepsy.

The state is directly visible to the observer in a Markov model. In a hidden Markov model (HMM), the state is not directly visible, but some outcome, dependent on the state, is visible [6, 35]. HMMs can be applied to many domains and have turned out to be particularly useful in several biological contexts. For example, this model structure has been applied for modelling the individual epileptic

M. Lavielle
Inria & Ecole Polytechnique, Université Paris-Saclay, France
E-mail: Marc.Lavielle@inria.fr

activity [1,9,24,38]. Indeed, individual profiles of epileptic seizures usually show distinct periods of low and high epileptic activity. Then, observed seizure frequencies would represent the observed output, while the unobserved state would represent the disease activity.

HMMs have shown to be helpful when characterizing different stages of migraines from a headache score (observed variable) [3,26]. Transitions between specific sleep states have also been successfully described using a hidden Markov model [11].

In the HMMs considered above, the state space of the hidden variables is discrete with a small number of states, while the observations themselves can either be discrete (typically generated from a categorical or a count distribution) or continuous (typically from a Gaussian distribution).

Hidden Markov models can also be generalized to allow continuous state spaces. In other words, the sequence of latent variables is a discrete-time Markovian process that can take any value in some continuous set. In particular, a Gaussian autoregressive process of order 1 is a Markovian process with memory 1 that can take any real values. The autoregressive model of order 1, AR(1), specifies that the prediction of the next value depends linearly on the current value [19].

Markov chains and autoregressive models are discrete-time processes. A value of the variable is defined each hour, day, week, On the other hand, biological phenomena are continuous-time phenomena. Then, the value of any biological parameter should be well defined at any time. Discrete-time processes implicitly assume that this parameter value is piecewise constant (during one hour, one day, one week, . . .) and suddenly jumps at the end of each time interval.

Such an assumption is acceptable if the latent variable is a hidden state, such as low and high epileptic activity, for instance, but probably not if this latent variable is a PK parameter such as the absorption rate constant or the clearance. If these PK characteristics are now considered as continuous time varying functions, then, continuous processes should be used for properly modelling the fluctuations of these parameters over time.

This naturally leads us to introduce a stochastic differential equations (SDEs) based model as an extension of an ordinary differential equations (ODEs) based model [30]. In particular, we show that the continuous-time Ornstein–Uhlenbeck diffusion process is the limit of a discrete-time AR(1) process when the time interval tends to 0. Then, a diffusion model is a continuous-time and continuous-state space Markovian process well-suited for properly describing biological phenomena. The use of such stochastic process has therefore been widely used for PKPD model development [5,16,17,22,25,33,36].

In the population approach, mixed effects hidden Markov models for discrete time processes and mixed effects diffusion models for continuous time processes aims to describe each individual series of observations using either a HMM or a system of SDEs while also taking into account variability between individuals [2,8,10,12,13,21,25].

Parameter estimation in mixed effects models with (hidden) Markovian dynamics then require specific methodologies. New inference procedures for mixed effects HMMs were proposed in [7]. These methods combine the SAEM algorithm with the Forward/Backward algorithm for maximum likelihood estimation of the population parameters and the Viterbi algorithm for recovering the individual sequences of hidden states [6]. Note that these methods are now available in NONMEM [4].

Donnet and Samson propose a survey of existing estimation methods for PKPD models based on SDEs [15]. A first comment is that most parametric estimation methods proposed for SDEs require high frequency data and are often poorly suited for sparse PKPD data.

In simple SDE models, such as a linear dynamical system, exact inference is tractable using the Kalman recursive filter for computing the exact likelihood [16,18]; however, in general, exact inference in diffusion models is infeasible, and approximate methods must be used, such as the extended

Kalman filter. Specific combinations of the extended Kalman filter with the FOCE method have been suggested in [21,25,31,37], and with the SAEM algorithm in [8].

Approximations of the likelihood using the particle filter were proposed in [14,40]. Other approximations of the likelihood in mixed-effects diffusion models can be found in [32,34].

The paper is organized as follows. Section 2 is devoted to hidden Markov models. We use a Poisson model to illustrate the differences between a mixture model, where the sequence of states over time are independent random variables, and a Poisson HMM where the states form a Markov chain. An application to epileptic seizures counts illustrates the extension to mixed effects HMM. Continuous state-space Markovian models, including autoregressive models and diffusion models are introduced in Section 3. A very simple PK model with time varying clearance illustrates these different models. An application to the theophyllin PK data shows that random fluctuations of this PK parameter may explain a significant part of the within subject variability of the data.

2 Hidden Markov Models

2.1 The individual approach

2.1.1 The model

HMMs were developed to describe how a given system moves from one state to another over time in situations where the successive visited states are unknown and a set of observations is the only available information to describe the system's dynamics [6]. HMMs can be seen as a variant of mixture models that allow for possible memory in the sequence of hidden states.

An HMM is thus defined as a pair of processes $((z_j, y_j), j = 1, 2, \dots)$, where the latent sequence (z_j) is a Markov chain and the distribution of observation y_j at time t_j depends on state z_j (see Figure 1).

Here, we consider a parametric framework with homogeneous Markov chains on a discrete and finite state space $\{1, 2, \dots, L\}$. Then, there exists transition probabilities $(p_{k\ell}, 1 \leq k, \ell \leq L)$ such that, for all $j = 1, 2, \dots, n$,

$$\mathbb{P}(z_j = \ell | z_{j-1} = k) = p_{k\ell} \quad ; \quad 1 \leq k, \ell \leq L$$

and where $\sum_{\ell=1}^L p_{k\ell} = 1$ for all $k \in \{1, 2, \dots, L\}$.

The joint distribution of (z_1, \dots, z_n) is then defined by the transition matrix $P = (p_{k\ell}, 1 \leq k, \ell \leq L)$ and the probability distribution of the initial state z_0 .

There are several possible options for defining the distribution of z_0 :

1. a first option consists in assuming that z_0 is uniformly distributed on $\{1, 2, \dots, L\}$:

$$\mathbb{P}(z_0 = \ell) = 1/L \quad ; \quad 1 \leq \ell \leq L$$

2. assuming that the sequence $(z_j, j \geq 0)$ is stationary means that there exists a vector of probabilities $\pi = (\pi_1, \pi_2, \dots, \pi_L)$ such that $\mathbb{P}(z_j = \ell) = \pi_\ell$ for any $\ell \in \{1, 2, \dots, L\}$ and any $j \geq 0$. π is therefore the distribution of z_0 and is solution of

$$\pi = \pi P.$$

3. the simplest option consists in fixing z_0 to some arbitrary value in $\{1, 2, \dots, L\}$. For instance, we can select the most likely value, i.e. the value ℓ which maximizes $\pi_\ell = \mathbb{P}(z_j = \ell)$.

The choice of the initial distribution has usually a very limited impact on the results in practice. A uniform distribution will be used in the numerical examples presented below but any other choice would also be valid and lead to (almost) the same results.

Conditionally on $z_j = \ell$ the distribution of the observation y_j is assumed to be a parametric distribution whose parameter depends on the state ℓ . Poisson distributions will be used in the example below for modelling daily counts of epileptic seizures. In this model, the parameters are therefore the Poisson rates $(\lambda_\ell, 1 \leq \ell \leq L)$ where

$$y_j | z_j = \ell \sim \text{Poisson}(\lambda_\ell)$$

Parameters of a Poisson hidden Markov model are $\psi = (P, \lambda_1, \dots, \lambda_L)$ where P is the transition matrix of the Markov chain.

Remark: A *mixture model* can be seen as an extension of a HMM, without any memory. Indeed, a mixture model assumes that the states are independent, i.e. $\mathbb{P}(z_j = \ell | z_{j-1} = k) = \mathbb{P}(z_j = \ell) = \pi_\ell$. Then, the observations are also independent and

$$y_j \underset{\text{i.i.d.}}{\sim} \sum_{\ell=1}^L \pi_\ell \text{Poisson}(\lambda_\ell)$$

Parameters of a Poisson mixture model are $\psi = (\pi_1, \dots, \pi_L, \lambda_1, \dots, \lambda_L)$

2.1.2 Inference in HMM

Infering hidden Markov models is challenging, mostly due to the complex expression of the likelihood and to the non observable visited states. As a consequence, hidden Markov models are associated with three basic problems:

1. Compute the likelihood function, i.e. the probability distribution function of the observations $\mathcal{L}(\psi; y) = \mathbf{p}(y; \psi)$. The *Forward-Backward algorithm* (FB) is a recursive procedure for solving this problem with a limited complexity.
2. Estimate the vector of model parameters ψ . The maximum likelihood estimate $\hat{\psi}$ maximizes the likelihood $\mathcal{L}(\psi; y)$. The Expectation-Maximization (EM) algorithm for HMM is known as the *Baum-Welch algorithm* and can be used for computing $\hat{\psi}$.
3. Estimate the unknown sequence of states (z_j) . The most probable sequence of states is the one that maximizes the conditional distribution $\mathbb{P}(z_1, z_2, \dots, z_n | y_1, \dots, y_n)$. This sequence can be computed using the *Viterbi algorithm*.

These algorithms are referenced and discussed in [35]. They are implemented in most of the packages for inference in HMM, including the `mhsmm` R package [29] that we will use in the next section.

2.1.3 Example

We will use the data previously analyzed by Delattre et al. [9]. The data displayed in Figure 2 consists of daily counts of epileptic seizures measured over 237 days for a representative patient of the study.

A first model proposed Trokoñiz et al. in [38] assumes that the daily number of seizures are independent Poisson random variables with intensity λ .

Here, the maximum likelihood estimate of λ is $\bar{y} = 6.12$. We clearly see from Figure 3(a) that the estimated Poisson distribution doesn't fit well the empirical distribution of the data. Indeed the overdispersion in the data is not taken into account in this model since λ is both the mean and the variance of the number of daily seizures in a basic Poisson model [38].

In order to better take into account this overdispersion, the authors proposed a mixture of two Poisson distributions with parameters λ_1 and λ_2 . The estimated parameters of the mixture obtained with the EM algorithm were $\hat{\pi}_1 = 0.84$, $\hat{\pi}_2 = 0.16$, $\hat{\lambda}_1 = 3.59$ and $\hat{\lambda}_2 = 19.56$. Figure 3(b) shows that the fit is improved with this mixture model.

Finally, we see from Figure 3(c) that the empirical distribution of the data is very well described by a mixture of three Poisson distributions with the following parameters obtained with the EM algorithm: $\hat{\pi}_1 = 0.749$, $\hat{\pi}_2 = 0.200$, $\hat{\pi}_3 = 0.051$, $\hat{\lambda}_1 = 3.07$, $\hat{\lambda}_2 = 10.99$ and $\hat{\lambda}_3 = 32.11$.

Once the vector of model parameters ψ has been estimated, each state z_j , $1 \leq j \leq n$, is estimated by maximizing the conditional distribution $\mathbb{P}(z_j = \ell | y_j, \hat{\psi})$. The estimated states are displayed in Figure 4.

However, according to Delattre et al. [9], it is not realistic to assume that the states are independent if these states correspond to different levels of epileptic activity. Indeed, epileptic patients are reasonably expected to stay in the same state several days rather than switching randomly every day between states. Delattre et al. then propose to describe the dynamics of the states by using a discrete Markov model

The Baum-Welch algorithm implemented in the `mhsmm` R package was used to find the unknown parameters of the model, assuming two and three states:

$$\begin{aligned} \mathcal{M}_{2\text{states}} : \quad \hat{\lambda} &= \begin{pmatrix} 3.5 \\ 18.9 \end{pmatrix} & \hat{P} &= \begin{pmatrix} 0.890 & 0.110 \\ 0.533 & 0.467 \end{pmatrix} \\ \mathcal{M}_{3\text{states}} : \quad \hat{\lambda} &= \begin{pmatrix} 3.1 \\ 11.2 \\ 32.2 \end{pmatrix} & \hat{P} &= \begin{pmatrix} 0.875 & 0.096 & 0.029 \\ 0.355 & 0.542 & 0.103 \\ 0.527 & 0.305 & 0.168 \end{pmatrix} \end{aligned}$$

The states (z_j) estimated using the `mhsmm` package are displayed in Figure 5. It is worth to highlight some interesting differences between the solutions obtained from a mixture model and a HMM. If we consider the model with 3 states, a mixture model will predict $\hat{z}_j = 1$ if $y_j \leq 6$, $\hat{z}_j = 2$ if $7 \leq y_j \leq 20$ and $\hat{z}_j = 3$ if $y_j \geq 21$. On the other hand, we can see around $t = 50$ days or $t = 150$ days that the classification rule of a HMM takes into account not only the value of the observed count, but also the estimated states the day before and the day after. In particular, even if few seizures have been observed during days 49 and 151, these two days are classified in state 2 rather than in state 1 because the neighboring days are also classified in state 2.

2.2 The population approach

2.2.1 Mixed effects hidden Markov model

A mixed effects hidden Markov model (MEHMM) is a straightforward extension of an HMM. In a population context, an HMM is first specified for each individual. The N individual HMMs are

supposed independent with the same structure, but their parameters may vary from one individual to another.

Let ψ_i denote the set of parameters of the i^{th} HMM, $i = 1, \dots, N$. We have seen in Section 2.1.1 that the vector of parameters ψ_i of a Poisson HMM is basically composed of the transition probabilities ($p_{k\ell,i} = \mathbb{P}(z_{i,j} = \ell | z_{i,j-1} = k)$) and the Poisson intensities $(\lambda_{\ell,i})$.

The *population distribution* of the parameters is the probability distribution of the vector of individual parameters ψ_i :

$$\psi_i \sim \mathbf{p}(\cdot; \theta)$$

where θ is the vector of *population parameters*.

2.2.2 Inference in mixed effects HMM

Usual tasks, methods and algorithms for mixed effects models include

1. Estimation of the population parameters θ .

The maximum likelihood estimate $\hat{\theta}$ maximizes the observed likelihood $\mathcal{L}(\theta; y)$. EM-like algorithms (SAEM, MCEM, MCPEM [23]) as well as approximation based algorithms (FO, FOCE, Laplace [39]) require to be able to compute, for each individual i , the *complete likelihood*

$$\begin{aligned} \mathcal{L}(\theta; y_i, \psi_i) &= \mathbf{p}(y_i, \psi_i; \theta) \\ &= \mathbf{p}(y_i | \psi_i) \mathbf{p}(\psi_i; \theta) \end{aligned}$$

The population pdf $\mathbf{p}(\psi_i; \theta)$ of the individual parameters usually has a closed-form expression (when the random effects are normally distributed for instance). On the other hand, we have seen in Section 2.1.2 that the Forward-Backward algorithm can be used for computing the conditional pdf $\mathbf{p}(y_i | \psi_i)$.

All these estimation methods can therefore be combined with the FB algorithm for estimating θ .

2. Computation of the likelihood function

$$\begin{aligned} \mathcal{L}(\hat{\theta}; y) &= \mathbf{p}(y; \hat{\theta}) \\ &= \sum_{i=1}^N \int \mathbf{p}(y_i | \psi_i) \mathbf{p}(\psi_i; \hat{\theta}) d\psi_i \end{aligned}$$

Monte Carlo integration methods (e.g. Importance Sampling) as well as approximation based methods (e.g. linearization, Laplace) can then be combined with the FB algorithm for evaluating the likelihood function $\mathcal{L}(\hat{\theta}; y)$.

3. Estimation of the individual parameters $(\psi_i, 1 \leq i \leq N)$.

For $i = 1, 2, \dots, N$, the empirical Bayes estimate (EBE) $\hat{\psi}_i$ maximizes the conditional distribution

$$\mathbf{p}(\psi_i | y_i; \hat{\theta}) = \frac{\mathbf{p}(y_i | \psi_i) \mathbf{p}(\psi_i; \hat{\theta})}{\mathbf{p}(y_i; \hat{\theta})}$$

Here again, the product $\mathbf{p}(y_i | \psi_i) \mathbf{p}(\psi_i; \hat{\theta})$ can be maximized as soon as the conditional pdf $\mathbf{p}(y_i | \psi_i)$ can be computed using the FB algorithm.

4. For each individual, estimate the unknown sequence of states $(z_{i,j}, 1 \leq j \leq n_i)$.

Once the vector of individual parameter ψ_i has been estimated, the Viterbi algorithm introduced in Section 2.1.2 allows one to obtain the most probable sequence of states $(\hat{z}_{i,1}, \dots, \hat{z}_{i,n_i})$ that maximizes the conditional distribution $\mathbf{p}(z_{i,1}, \dots, z_{i,n_i} | y_{i,1}, \dots, y_{i,n_i}, \hat{\psi}_i)$.

2.2.3 Application to epileptic seizure count data

The data base analyzed by Delattre et al. in [9] consisted of six clinical studies. Recruited patients were on standard anti-epileptic therapy during several weeks before being randomized to parallel treatment groups. Delattre et al. proposed to describe exposure-response relationship of an antiepileptic agent, using a mixed effects HMM. This analysis shown demonstrated that MEHMM is able to mimic dynamics of seizure frequencies very well.

A complete model was developed for both the screening and the active treatment phases in [9] but we will restrict our analysis to the screening phase. The data therefore consists of 71447 daily counts of epileptic seizures measured between 3 and 30 weeks for $N = 788$ patients.

An homogeneous Markov chain with two states is used for these data:

$$\begin{aligned}\mathbb{P}(z_{i,j} = 1 | z_{i,j-1} = 1) &= 1 - \mathbb{P}(z_{i,j} = 2 | z_{i,j-1} = 1) = p_{11,i} \\ \mathbb{P}(z_{i,j} = 2 | z_{i,j-1} = 2) &= 1 - \mathbb{P}(z_{i,j} = 1 | z_{i,j-1} = 2) = p_{22,i}\end{aligned}$$

Then, the distribution of the seizure counts is a mixture of two Poisson distributions:

$$\begin{aligned}y_{i,j} | z_{i,j} = 1 &\sim \text{Poisson}(\lambda_{1,i}) \\ y_{i,j} | z_{i,j} = 2 &\sim \text{Poisson}(\lambda_{2,i})\end{aligned}$$

According to this model, the vector of individual parameters for patient i is $\psi_i = (p_{11,i}, p_{22,i}, \lambda_{1,i}, \lambda_{2,i})$. We assume logit-normal distributions for $p_{11,i}$ and $p_{22,i}$ and log-normal distributions for $\lambda_{1,i}$ and $\lambda_{2,i}$:

$$\begin{aligned}\text{logit}(p_{11,i}) &\sim \mathcal{N}(\text{logit}(p_{11,\text{pop}}), \omega_{11}^2) \\ \text{logit}(p_{22,i}) &\sim \mathcal{N}(\text{logit}(p_{22,\text{pop}}), \omega_{22}^2) \\ \log(\lambda_{1,i}) &\sim \mathcal{N}(\log(\lambda_{1,\text{pop}}), \omega_1^2) \\ \log(\lambda_{2,i}) &\sim \mathcal{N}(\log(\lambda_{2,\text{pop}}), \omega_2^2)\end{aligned}$$

where $\text{logit}(x) = \log(x/(1-x))$ for $0 < x < 1$.

Here, the vector of population parameters is $\theta = (p_{11,\text{pop}}, p_{22,\text{pop}}, \lambda_{1,\text{pop}}, \lambda_{2,\text{pop}}, \omega_{11}^2, \omega_{22}^2, \omega_1^2, \omega_2^2)$

Using the proposed model, we can estimate what is, for each individual, the individual probability of transitioning to the high/low activity state as well as the probability of remaining in the same state.

Estimation of the population parameter θ , the individual parameters (ψ_i) and the individual sequences of states (z_{ij}) were performed with the Monolix software (version 4.3.3) combined with a Matlab implementation of the Forward-Backward and Viterbi algorithms. The Monolix project and the Matlab codes are available as supplementary material.

Results obtained with a mixture model and a HMM are presented in Table 1.

The observed data and the estimated states for three representative patients are displayed in Figure 6. Huge inter individual variability of the distribution of the seizure counts can be clearly seen in this Figure. Nevertheless, a Markovian dynamics between two states of epileptic activity seems to properly describe the data for these three patients (similar results are obtained with the other patients of the study).

3 Continuous State-Space Markovian Models

3.1 The individual approach

3.1.1 Discrete time model

We were assuming in the previous section that there exists some latent variable that takes a finite number of values. Such an assumption may be a good approximation when this latent variable aims to describe the state of a patient over time, for instance, but it may be not appropriated when the phenomena to describe cannot be reduced to a finite number of states.

Consider, for instance, the pharmacokinetics of a drug which characteristics may change over time. Then, it is not realistic to assume a finite number of PK characteristics for a given patient (slow/fast absorption, slow/fast elimination, ...).

A widely used approach consists of splitting the study into M time periods, or “occasions” and assuming that some PK parameters can vary from occasion to occasion but remain constant within occasion. In other word, the sequence of PK parameters $(\psi_m, 1 \leq m \leq M)$ randomly fluctuates around some typical vector of PK parameters ψ^* . In pharmacometrics, an occasion usually represents an entire dosing cycle, i.e. the time interval between two successive doses, but this definition can easily be extended to any arbitrary sequence of time intervals.

As an illustration, consider a one compartment PK model for a single bolus administration:

$$A(t) = De^{-\frac{Cl}{V}t}$$

where D is the amount of drug administered at time $t = 0$. Here, $A(t)$ is the amount of drug in the central compartment at time t , V the volume of the central compartment and Cl is the drug clearance.

A simple model for *inter occasion variability* (IOV) assumes that Cl_m , the clearance during occasion m , is log-normally distributed around some typical value Cl^*

$$\log(Cl_m) = \log(Cl^*) + \zeta_m \quad ; \quad m = 1, 2, \dots, M \quad (1)$$

where $\zeta_m \sim_{\text{i.i.d.}} \mathcal{N}(0, \tau^2)$.

Let δ be the length of an occasion and let $t_m = m\delta$ for $m = 0, 1, \dots, M$. Then, $Cl(t) = Cl_m$ for $t_{m-1}\delta \leq t < t_m\delta$ and, conditionally to the sequence (Cl_m) , the predicted amount defined by the model is

$$\begin{aligned} A(0) &= D \\ A(t) &= A(t_{m-1})e^{-\frac{Cl_m}{V}(t-t_{m-1})} \quad ; \quad t_{m-1} \leq t < t_m \end{aligned}$$

An example is displayed in Figure 7. There are only three occasions of length $\delta = 10$ h in this example. Two sequences $(Cl_m, 1 \leq m \leq 3)$ were simulated with $\tau = 0.20$ and $\tau = 0.50$, respectively, $Cl^* = 3 \text{ L.h}^{-1}$, $V = 1 \text{ L}$ and $D = 1 \text{ g}$.

In this model, the sequence (Cl_m) of clearances is independent. Then, the model allows any jump between slow and fast eliminations. If we prefer to assume that the clearance cannot drastically change from one day to the next, a Markovian dynamics can be introduced in order to better control the fluctuations of this parameter around Cl^* .

Let

$$X_m = \log(Cl_m) - \log(Cl^*)$$

be the deviation of the log-clearance from its typical value at period m . A discrete-time Markov process with memory 1 may be represented by an *autoregressive process* of order 1 (AR(1) process).

$$X_{m+1} = \rho X_m + \zeta_{m+1} \quad (2)$$

where $0 \leq \rho < 1$ and where $\zeta_m \sim_{\text{i.i.d.}} \mathcal{N}(0, \tau^2)$. Furthermore, ζ_{m+1} is independent of $(X_\ell, \ell \leq m)$ in this model. Because of this property, the sequence (ζ_m) is known as the *innovation* of the process X .

We can easily check that ρ is the correlation between X_m and X_{m+1} . More generally, ρ^ℓ is the correlation between X_m and $X_{m+\ell}$.

When $\rho = 0$, the X_m 's are independent and identically distributed around 0. The model is indeed the one represented in equation (1): the Cl_m are independent and log-normally distributed around the typical value Cl^* .

On the other hand, when ρ is close to 1, X_{m+1} tends to be close to the previous value X_m , which also means that Cl_{m+1} tends to be close to Cl_m . More precisely, $(Cl_m, m \geq 1)$ is a Markov chain with a log-Gaussian transition kernel

$$\log(Cl_{m+1})|Cl_m \sim \mathcal{N}(\rho \log(Cl_m) + (1 - \rho) \log(Cl^*), \tau^2)$$

$\delta = 1$ h was used in the example displayed in Figure 8. We can see that when both the variance τ^2 and the correlation ρ are small ($\tau^2 = 0.25^2, \rho = 0.1$) the predicted amount profile remains very close to the typical amount profile obtained with $Cl^* = 3 \text{ L.h}^{-1}$. On the other hand, successive stages of slow and fast elimination clearly appear with higher values for these parameters ($\tau^2 = 0.50^2, \rho = 0.8$).

3.1.2 Stochastic differential equation based model

The AR model represented in (2) assumes that some PK parameter is piecewise constant and suddenly jumps at the end of each period. Such an assumption may not be biologically realistic and a continuous process should probably be considered for modelling the fluctuations of this parameter over time.

One approach consists in letting the period length δ tend to 0. Since the limiting process (when $\delta = 0$) is a continuous process, the correlation ρ of the process should tend to 1 and the variance τ^2 of the innovation should tend to 0 when δ goes to 0. These constraints are satisfied setting, for instance, $\rho = (1 - \alpha\delta)$ and $\tau = \gamma\sqrt{\delta}$ where $0 \leq \alpha < 1$ and $\gamma > 0$. With this new parameterization, the AR model now writes

$$X_{m+1} = (1 - \alpha\delta)X_m + \gamma\sqrt{\delta}e_{m+1}$$

or, equivalently,

$$X_{m+1} - X_m = -\alpha X_m \delta + \gamma\sqrt{\delta}e_{m+1}$$

where $e_{m+1} = \zeta_{m+1}/\tau \sim \mathcal{N}(0, 1)$. Letting δ go to 0 leads to the *Ornstein-Uhlenbeck diffusion process*

$$dX(t) = -\alpha X(t)dt + \gamma dW(t)$$

where $(W(t), t \geq 0)$ is a standard Wiener process in \mathbb{R} :

$$\begin{aligned} W(t) &\sim \mathcal{N}(0, t) \\ \mathbb{E}(W(t)(W(t+s) - W(t))) &= 0 \end{aligned}$$

We can extend the discrete-time model for the clearance Cl presented in Section 3.1.1 to this continuous-time model:

$$d \log(Cl(t)) = -\alpha (\log(Cl(t)) - \log(Cl^*)) dt + \gamma dW(t)$$

$(Cl(t), t \geq 0)$ is now a continuous-time Markovian process which takes its values in \mathbb{R}^+ . The PK model combines this diffusion model for $Cl(t)$ with an ODE model for the amount $A(t)$:

$$\begin{aligned} \frac{d \log(Cl(t))}{dt} &= -\alpha (\log(Cl(t)) - \log(Cl^*)) + \gamma \frac{dW(t)}{dt} \\ \frac{dA(t)}{dt} &= -\frac{Cl(t)}{V} A(t) \end{aligned} \quad (3)$$

where $A(0) = D$.

The numerical example displayed in Figure 9 shows that when the coefficient of the Wiener process term is small ($\gamma = 1.5$) and the coefficient of the drift is large ($\alpha = 10$), then $Cl(t)$ remains close to the typical value Cl^* and the predicted amount profile remains very close to the typical amount profile obtained with $Cl^* = 3 \text{ L.h}^{-1}$. Fluctuations show more randomness with $\gamma = 3.5$ and $\alpha = 5$.

PK model represented in (3) can easily be extended to more complex models. The absorption rate constant of a model for oral administration, for instance, could also be described by a diffusion model if we suspect that the rate of absorption may fluctuate:

$$\begin{aligned} \frac{d \log(ka(t))}{dt} &= -\alpha_1 (\log(ka(t)) - \log(ka^*)) + \gamma_1 \frac{dW_1(t)}{dt} \\ \frac{d \log(Cl(t))}{dt} &= -\alpha_2 (\log(Cl(t)) - \log(Cl^*)) + \gamma_2 \frac{dW_2(t)}{dt} \\ \frac{dAd(t)}{dt} &= -ka(t)Ad(t) \\ \frac{dAc(t)}{dt} &= ka(t)Ad(t) - \frac{Cl(t)}{V} Ac(t) \end{aligned} \quad (4)$$

where $Ad(0) = D$ and $Ac(0) = 0$. Here, Ad and Ac are the amounts of drug in the depot and central compartments.

Both $(ka(t), t \geq 0)$ and $(Cl(t), t \geq 0)$ are Markov processes in this model.

It is also possible to consider that some parameters of the model randomly vary over time while other parameters remain constant. If, for example, the rate of absorption is constant over time, then the first equation of system (4) should be replaced by $ka(t) = ka$.

An Ornstein-Uhlenbeck diffusion process is very convenient for easily introducing continuous random fluctuations of some parameters around some typical values. There are also more general

stochastic differential equations that could be considered. For instance, the solution of the typical stochastic differential equation (SDE)

$$dX(t) = \mu(X(t), t)dt + \nu(X(t), t)dW(t)$$

is called a *diffusion process* and is a Markov process under general conditions on functions μ and ν [20]. Functions μ and ν are known, respectively, as the drift and the diffusion coefficient of X . In pharmacology, X can be a time varying PK parameter for instance.

There are also more general SDE's where the functions μ and ν depend not only on the present value of the process $X(t)$, but also on previous values of the process. In that case the defining equation is called a stochastic delay differential equation. If μ and/or ν depend on present or previous values of other processes, the solution process $X(t)$ is not a Markov process, and it is called an *Itô process* and not a diffusion process.

3.1.3 Inference in a diffusion model

We consider parametric diffusion models of the form

$$dX(t) = \mu(X(t), \psi)dt + \nu(X(t), \psi)dW(t) \quad (5)$$

where ψ is a vector of parameters.

In practice, X is an *hidden process* since it cannot be observed directly. Observations (y_1, y_2, \dots, y_n) are some noisy functions of $(X(t_1), \dots, X(t_n))$,

$$y_j = g(X(t_j), \psi) + \varepsilon_j$$

where $\varepsilon_j \sim_{\text{i.i.d.}} \mathcal{N}(0, \sigma^2)$.

PK model (3) for bolus administration can be written in this form with

$$X(t) = \begin{pmatrix} \log(Cl(t)) \\ A(t) \end{pmatrix} \quad ; \quad \mu(X(t)) = \begin{pmatrix} -\alpha(\log(Cl(t)) - \log(Cl^*)) \\ -\frac{e^{\log(Cl(t))}}{V}A(t) \end{pmatrix} \quad ; \quad \nu(X(t)) = \begin{pmatrix} \gamma \\ 0 \end{pmatrix}$$

In this model, observations are observed concentrations and $g(X(t)) = A(t)/V$.

Parameters of this model are therefore $\psi = (Cl^*, \alpha, \gamma, V, \sigma)$. Maximum likelihood estimation of ψ requires to compute the likelihood

$$\begin{aligned} \mathcal{L}(\psi; y) &= \mathbf{p}(y; \psi) \\ &= \mathbf{p}(y_1; \psi) \mathbf{p}(y_2 | y_1; \psi) \cdots \mathbf{p}(y_n | y_{n-1}, \dots, y_1; \psi) \end{aligned}$$

Except in some very specific classes of diffusion models, the transition density $\mathbf{p}(y_j | y_{j-1}, \dots, y_1; \psi)$ does not have a closed-form expression since it involves the transition densities of the underlying diffusion processes X . When the underlying system is a Gaussian linear dynamical one, this density is Gaussian with mean and variance that can be computed using the Kalman filter. When the system is not linear, a solution comes from approximating this density by a Gaussian one and using the extended Kalman filter (EKF) for quickly computing its mean and variance [8].

Once ψ has been estimated, it is possible to recover the latent process $(X(t), t \geq 0)$ at the measurement times (t_1, \dots, t_n) . The fixed-interval Kalman smoother [18] estimates $(X(t_1), \dots, X(t_n))$ as the most likely sequence of hidden variables:

$$(\hat{X}(t_1), \dots, \hat{X}(t_n)) = \underset{X(t_1), \dots, X(t_n)}{\operatorname{argmax}} \mathbf{p}(X(t_1), \dots, X(t_n) | y_1, \dots, y_n, \hat{\psi}).$$

Implementation of the EKF requires to solve a system of ordinary differential equations that generally do not have any closed form solution. A simplified version of the numerical resolution method of these ODEs and based on higher-order Taylor approximations has been proposed by Mazzoni [27].

We should remark that this method only requires the knowledge of the Jacobian function of the drift function μ of the dynamical process. It is therefore easily implementable for any diffusion model such that derivatives of μ can be computed. Note also that this approach can be easily extended to models involving multidimensional observations.

In comparison to the EKF, particle filters do not require making approximations of the transition density, but are quite demanding in terms of simulation requirements and computation time (see [40] for an application in quantitative biology).

3.2 The population approach

3.2.1 Mixed effects diffusion model

Diffusion model introduced above can be extended in a straightforward manner to *mixed effects diffusion model* (MEDM) by defining system's parameters as a vector of individual parameters.

In other words, N individual diffusion models are specified with the same structure as the one represented in (5), but their parameters ψ_i may vary from one individual to another.

For an Ornstein-Uhlenbeck diffusion process, ψ_i is basically composed of the coefficients α_i and γ_i and the typical parameter value ψ_i^* .

As usual, ψ_i is treated as a random parameter whose distribution depends on a vector of population parameters θ .

3.2.2 Inference in mixed effects diffusion models

Here again, the first tasks to perform consist in estimating the vector of population parameters θ , computing the likelihood function and estimating the individual parameters (ψ_i).

It was shown in Section 2.2.2 that suitable methods and algorithms for performing these tasks require to be able to compute the conditional pdf of the observations $p(y_i|\psi_i)$ for each individual i . The Forward-Backward algorithm was shown to be very efficient for computing this conditional pdf for HMM based model. In the case of a diffusion based model, we have seen in Section 3.1.3 that the extended Kalman filter can be used for computing $p(y_i|\psi_i)$.

The EKF combined with “standard” inference algorithms for mixed effects model (FOCE, SAEM, Importance Sampling, ...) can then be used for mixed effects diffusion models. For a combination of the FOCE approximation of the population likelihood and the EKF, see for example [21, 31, 37, 25]. The combination of the SAEM algorithm and EKF proposed in [8] is shown to have very good statistical properties. See [15] for a review of parameter estimation methods in SDE population models.

Once the population and individual parameters have been estimated, each individual hidden sequence $(X_i(t_j), 1 \leq j \leq n_i)$ can be recovered using the fixed-interval Kalman smoother [18] or the particle filter [14].

3.2.3 Application to the theophyllin data

The proposed approach was previously used in [8] to analyze the theophyllin PK data. These data come from twelve subjects who received a single oral dose of $D=320$ mg of theophyllin. For each subject, the concentration of theophyllin in plasma is measured through ten consecutive blood samples after the dose administration. These data are classically described by a one compartment model with first order absorption and linear elimination. For $i = 1, 2, \dots, N$,

$$\begin{aligned}\frac{dAd_i(t)}{dt} &= -k_{a_i}Ad_i(t) \\ \frac{dAc_i(t)}{dt} &= k_{a_i}Ad_i(t) - \frac{Cl_i}{V_i}Ac_i(t) \\ Cc_i(t) &= \frac{Ac_i(t)}{V_i}\end{aligned}\tag{6}$$

where $Ad_i(0) = D$ and $Ac_i(0) = 0$. Here, Cc is the concentration in the central compartment.

Following Delattre and Lavielle, we can assume that the clearance Cl_i is a diffusion process:

$$\begin{aligned}\frac{d\log(Cl_i(t))}{dt} &= -\alpha(\log(Cl_i(t)) - \log(Cl_i^*)) + \gamma\frac{dW_i(t)}{dt} \\ \frac{dAd_i(t)}{dt} &= -k_{a_i}Ad_i(t) \\ \frac{dAc_i(t)}{dt} &= k_{a_i}Ad_i(t) - \frac{Cl_i(t)}{V_i}Ac_i(t)\end{aligned}\tag{7}$$

Log-normal distributions describe the inter individual variability of the PK parameters:

$$\begin{aligned}\log(k_{a_i}) &\sim \mathcal{N}(\log(k_{a_{\text{pop}}}), \omega_{ka}^2) \\ \log(V_i) &\sim \mathcal{N}(\log(V_{\text{pop}}), \omega_{ka}^2) \\ \log(Cl_i^*) &\sim \mathcal{N}(\log(Cl_{\text{pop}}^*), \omega_{Cl}^2)\end{aligned}\tag{8}$$

k_{a_i} , V_i and Cl_i^* . On the other hand, coefficients α and γ of the diffusion are the same for the twelve subjects.

For each individual i , observed concentrations $(y_{ij}, 1 \leq j \leq n_i)$ measured at times $(t_{ij}, 1 \leq j \leq n_i)$ are assumed to be log-normally distributed around the concentrations predicted by the PK model:

$$y_{ij} = \frac{Ac_i(t_{ij})}{V_i} e^{\varepsilon_{ij}} \quad ; \quad 1 \leq j \leq n_i$$

where $\varepsilon_{ij} \sim_{\text{i.i.d.}} \mathcal{N}(0, \sigma^2)$

Monolix 4.3.3 was combined with a Matlab implementation of the EKF for estimating the parameters of the model.

When the solution of a system of SDEs is only observed at few time points with noise, it becomes difficult to distinguish stochastic variability from residual variability [8, 14, 34]. Without any prior information on the parameter values, maximum likelihood estimation of the model parameters leads to the solution obtained with the ODE based model ($\hat{\sigma} = 0.173$ with the theophyllin data). It is then necessary to introduce some informative prior in order to enforce the solution to include a significant stochastic component and not only residual variability.

Maximum likelihood estimation of the population PK parameters was therefore combined with a Bayesian estimation of α , γ and σ using log-normal priors:

$$\log(\alpha) \sim \mathcal{N}(\log(0.5), 0.05^2) \quad ; \quad \log(\gamma) \sim \mathcal{N}(\log(0.4), 0.05^2) \quad ; \quad \log(\sigma) \sim \mathcal{N}(\log(0.04), 0.05^2)$$

The Maximum a posteriori (MAP) estimate of these parameters was computed by maximizing the posterior distribution of α , γ and σ . Estimated parameters are given in Table 2. The Monolix project files and the Matlab implementation of the extended Kalman filter used for this example are available as supplementary material.

The predicted PK profiles for three subjects obtained with the two models are displayed in Figure 10. The predicted curves obtained under both models are close to the observed concentrations for the three subjects, but we see that the SDE based model leads to a much better fit since the model enables to capture smooth random variations of the concentration kinetics. We can also remark that these three patients share the same profile: a large value of the clearance during the first 6-7 hours indicate a first stage of fast elimination of the drug, followed by a slower elimination. Then, assuming that the elimination rate of theophyllin is not constant over time is a plausible hypothesis. Other models able to take into account this behavior should therefore be tested, including nonlinear elimination, peripheral compartment(s), or inter occasion variability.

4 Conclusions

Mixed effects hidden Markov models and mixed effects diffusion models have recently been defined as the extensions of HMM and SDE based models to population studies. They have shown to be relevant for modelling biological phenomena that involve some non observed Markovian dynamics.

Sound methodologies for these models have been developed and published during the last decade. Furthermore, several simulation based studies have confirmed their very good statistical behavior: estimators are consistent and show small biases.

These methods consists in treating the hidden states as “nuisance parameters” Indeed, both the Forward-Backward algorithm for HMM and the (extended) Kalman filter for diffusion models allows one to “ignore” these hidden states by integrating the joint distribution of the observations and the hidden process over the hidden process. Standard methods for inference in mixed effects models can then be efficiently combined with these algorithms. In particular, combining these algorithms with the SAEM algorithm allows for the computational burden to be minimal.

Applications on real data examples have demonstrated that these methodologies are also useful in practice, provided that rich data are available. On one hand, the use of mixed effects HMM offers a new insight in the analysis of epilepsy data where a better description of the individual epileptic activity can be taken into account as well as the variability between individuals. On the other hand, mixed effects diffusion models can be used as a realistic alternative to inter occasion variability for properly modelling within and between subject variability of PK parameters.

Unfortunately, implementation and use of MEHMM and MEDM remain quite limited and mainly - not to say only - in the realm of academic research. Indeed, these models cannot be easily implemented and used with the standard software tools available today for mixed effects modelling for pharmacometrics. Then, comparing, for instance, a classical IOV model with a SDE based model represents a tremendous challenge for a pharmacometrician in term of coding. Indeed, in the absence of reliable and easy-to-use tools, a strong statistical expertise is required for implementing the EKF. Having the possibility to easily implement such models in a software tool would probably help to popularize their use.

Further improvements of the present applications of MEHMM and MEDM deserve to be considered for future works. A question of interest concerns the general nature of the latent process. The main assumption for the analysis of seizure counts is that the dynamics of the epileptic activity is described with a two-state first-order Markov chain. We could also consider models with more than two states or higher-order Markov dependence. Other distributions than Poisson distribution could also be considered for the count data. An Ornstein-Uhlenbeck diffusion model was used for modelling the fluctuations of the clearance. Other SDE based models could also be used for the same purpose. Then, specific tools should be developed for model assessment and model selection.

References

1. Albert, P.S.: A two state Markov mixture model for a time series of epileptic seizure counts. *Biometrics* **47**(4), 1371–1381 (1991)
2. Altman, R.M.: Mixed hidden Markov models: an extension of the hidden Markov model to the longitudinal data setting. *Journal of the American Statistical Association* **102**(477), 201–210 (2007)
3. Anisimov, V.V., Maas, H.J., Danhof, M., Della Pasqua, O.: Analysis of responses in migraine modelling using hidden Markov models. *Statistics in Medicine* **26**(22), 4163–4178 (2007)
4. Bauer, R.J.: Hidden Markov Model Analysis with NONMEM. https://nonmem.iconplc.com/nonmem/hmm/readme_hmm.pdf
5. Berglund, M., Sunnåker, M., Adiels, M., Jirstrand, M., Wennberg, B.: Investigations of a compartmental model for leucine kinetics using non-linear mixed effects models with ordinary and stochastic differential equations. *Mathematical Medicine and Biology* p. dqr021 (2011)
6. Cappé, O., Moulines, E., Rydén, T.: *Inference in Hidden Markov Models*. Springer (2005)
7. Delattre, M., Lavielle, M.: Maximum likelihood estimation in discrete mixed hidden Markov models using the SAEM algorithm. *Computational Statistics & Data Analysis* **56**(6), 2073–2085 (2012)
8. Delattre, M., Lavielle, M.: Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Statistics and Its Interfaces* **6**(4), 519–532 (2013)
9. Delattre, M., Savic, R.M., Miller, R., Karlsson, M.O., Lavielle, M.: Analysis of exposure–response of CI-945 in patients with epilepsy: application of novel mixed hidden Markov modeling methodology. *Journal of Pharmacokinetics and Pharmacodynamics* **39**(3), 263–271 (2012)
10. Deng, C., Plan, E.L., Karlsson, M.O.: Approaches for modeling within subject variability in pharmacometric count data analysis: dynamic inter-occasion variability and stochastic differential equations. *Journal of pharmacokinetics and pharmacodynamics* **43**(3), 305–314 (2016)
11. Diack, C., Ackaert, O., Ploeger, B., van der Graaf, P., Gurrell, R., Ivarsson, M., Fairman, D.: A hidden Markov model to assess drug-induced sleep fragmentation in the telemetered rat. *Journal of pharmacokinetics and pharmacodynamics* **38**(6), 697–711 (2011)
12. Ditlevsen, S., De Gaetano, A.: Stochastic vs. deterministic uptake of dodecanedioic acid by isolated rat livers. *Bulletin of mathematical biology* **67**(3), 547–561 (2005)
13. Donnet, S., Samson, A.: Parametric inference for mixed models defined by stochastic differential equations. *ESAIM: Probability and Statistics* **12**, 196–218 (2008)
14. Donnet, S., Samson, A.: EM algorithm coupled with particle filter for maximum likelihood parameter estimation of stochastic differential mixed-effects models (2011). URL <https://hal.archives-ouvertes.fr/hal-00519576>. Working paper or preprint
15. Donnet, S., Samson, A.: A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models. *Advanced Drug Delivery Reviews* **65**(7), 929–939 (2013)
16. Favetto, B., Samson, A.: Parameter estimation for a bidimensional partially observed Ornstein–Uhlenbeck process with biological application. *Scandinavian Journal of Statistics* **37**(2), 200–220 (2010)
17. Ferrante, L., Bompadre, S., Leone, L.: A stochastic compartmental model with long lasting infusion. *Biometrical journal* **45**(2), 182–194 (2003)
18. Grewal, M.S., Andrews, A.P.: *Kalman Filtering: Theory and Practice Using MATLAB*. Wiley (2011)
19. Hamilton, J.D.: *Time series analysis*, vol. 2. Princeton university press Princeton (1994)
20. Itô, K.: *Diffusion processes*. Wiley Online Library (1974)
21. Klim, S., Mortensen, S.B., Kristensen, N.R., Overgaard, R.V., Madsen, H.: Population stochastic modelling (PSM) – an R package for mixed-effects models based on stochastic differential equations. *Computer Methods and Programs in Biomedicine* **94**(3), 279–289 (2009)
22. Kristensen, N.R., Madsen, H., Ingwersen, S.H.: Using stochastic differential equations for PK/PD model development. *Journal of pharmacokinetics and pharmacodynamics* **32**(1), 109–141 (2005)
23. Lavielle, M.: *Mixed effects models for the population approach: models, tasks, methods and tools*. CRC press (2014)
24. Le Cam, S., Louis-Dorr, V., Maillard, L.: Hidden Markov chain modeling for epileptic networks identification. In: *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, pp. 4354–4357. IEEE (2013)
25. Leander, J., Almquist, J., Ahlström, C., Gabrielsson, J., Jirstrand, M.: Mixed effects modeling using stochastic differential equations: illustrated by pharmacokinetic data of nicotinic acid in obese Zucker rats. *The AAPS journal* **17**(3), 586–596 (2015)
26. Maas, H., Danhof, M., Pasqua, O.D.: Prediction of headache response in migraine treatment. *Cephalalgia* **26**(4), 416–422 (2006)
27. Mazzoni, T.: Computational aspects of continuous–discrete extended Kalman-filtering. *Computational Statistics* **23**(4), 519–539 (2008)
28. Meyn, S., Tweedie, R.L.: *Markov Chains and Stochastic Stability*. Cambridge University Press (2009)
29. O’Connell, J., Højsgaard, S., et al.: Hidden semi Markov models for multiple observation sequences: The mhsmm package for R. *Journal of Statistical Software* **39**(4), 1–22 (2011)
30. Oksendal, B.: *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media (2013)

31. Overgaard, R., Jonsson, N., Tornøe, C., Madsen, H.: Non-linear mixed-effects models with stochastic differential equations: implementation of an estimation algorithm. *Journal of Pharmacokinetics and Pharmacodynamics* **32**(1), 85–107 (2005)
32. Picchini, U., Ditlevsen, S.: Practical estimation of high dimensional stochastic differential mixed-effects models. *Computational Statistics & Data Analysis* **55**(3), 1426–1444 (2011)
33. Picchini, U., Ditlevsen, S., De Gaetano, A.: Modeling the euglycemic hyperinsulinemic clamp by stochastic differential equations. *Journal of mathematical biology* **53**(5), 771–796 (2006)
34. Picchini, U., Gaetano, A., Ditlevsen, S.: Stochastic differential mixed-effects models. *Scandinavian Journal of Statistics* **37**(1), 67–90 (2010)
35. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **77**(2), 257–286 (1989)
36. Ramanathan, M.: An application of Ito's lemma in population pharmacokinetics and pharmacodynamics. *Pharmaceutical research* **16**(4), 584–586 (1999)
37. Tornøe, C.W., Overgaard, R.V., Agersø, H., Nielsen, H.A., Madsen, H., Jonsson, E.N.: Stochastic differential equations in NONMEM®: implementation, application, and comparison with ordinary differential equations. *Pharmaceutical research* **22**(8), 1247–1258 (2005)
38. Trocóniz, I.F., Plan, E.L., Miller, R., Karlsson, M.O.: Modelling overdispersion and Markovian features in count data. *Journal of pharmacokinetics and pharmacodynamics* **36**(5), 461 (2009)
39. Wang, Y.: Derivation of various nonmem estimation methods. *Journal of Pharmacokinetics and pharmacodynamics* **34**(5), 575–593 (2007)
40. Zechner, C., Unger, M., Pelet, S., Peter, M., Koeppl, H.: Scalable inference of heterogeneous reaction kinetics from pooled single-cell recordings. *Nature Methods* **11**, 197–202 (2014)

Table 1: Estimation of the population parameters for the epileptic seizure count data obtained with a mixture model and a hidden Markov model

Parameter	Mixture model		HMM model	
	Estimate	Stand. Err.	Estimate	Stand. Err.
$p_{1,\text{pop}}$	0.907	0.011	-	-
$p_{11,\text{pop}}$	-	-	0.87	0.007
$p_{22,\text{pop}}$	-	-	0.48	0.03
$\lambda_{1,\text{pop}}$	0.209	0.011	0.126	0.009
$\lambda_{2,\text{pop}}$	1.86	0.10	1.25	0.05
ω_{p_1}	1.94	0.10	-	-
$\omega_{p_{11}}$	-	-	0.88	0.05
$\omega_{p_{22}}$	-	-	2.17	0.12
ω_{λ_1}	1.23	0.04	1.44	0.05
$\omega_{\lambda_{21}}$	0.93	0.03	1.03	0.03

Table 2: Estimation of the population parameters for the theophyllin data with the ODE based model eq:pkode and the SDE based model eq:pkdsde3

Parameter	ODE model		SDE model	
	Estimate	Stand. Err.	Estimate	Stand. Err.
ka_{pop}	1.31	0.25	1.02	0.31
V_{pop}	32.2	1.6	25.2	3.5
Cl_{pop}^*	2.79	0.20	2.92	0.24
α	-	-	0.519 (MAP)	0.22
γ	-	-	0.433 (MAP)	0.11
ω_{ka}	0.627	0.160	0.968	0.240
ω_V	0.132	0.053	0.345	0.130
ω_{Cl^*}	0.239	0.056	0.164	0.094
σ	0.174	0.014	0.056 (MAP)	0.002

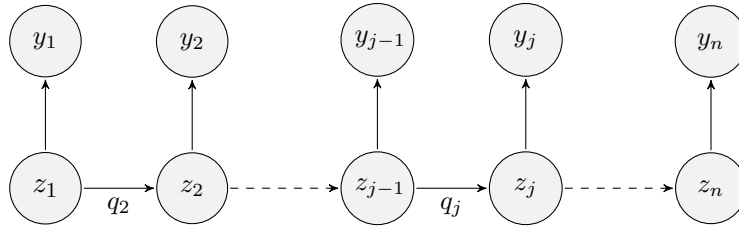


Fig. 1: Dynamics of a hidden Markov model.

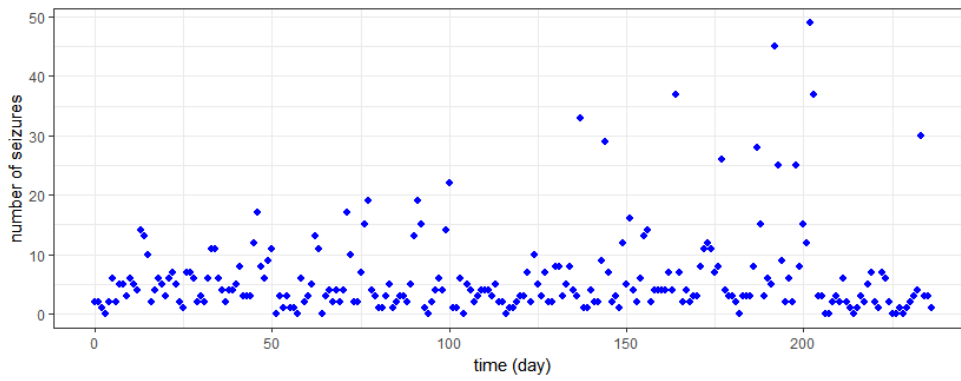


Fig. 2: Observed daily seizure counts for a single patient

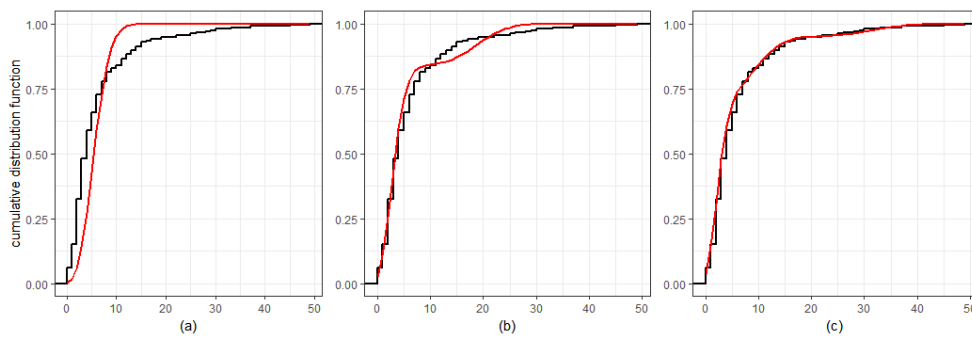


Fig. 3: Empirical cumulative distribution function of the sequence of the 237 daily seizure counts (black) and cdf of the fitted model (red): (a) Poisson distribution, (b) mixture of two Poisson distributions, (c) mixture of three Poisson distributions.

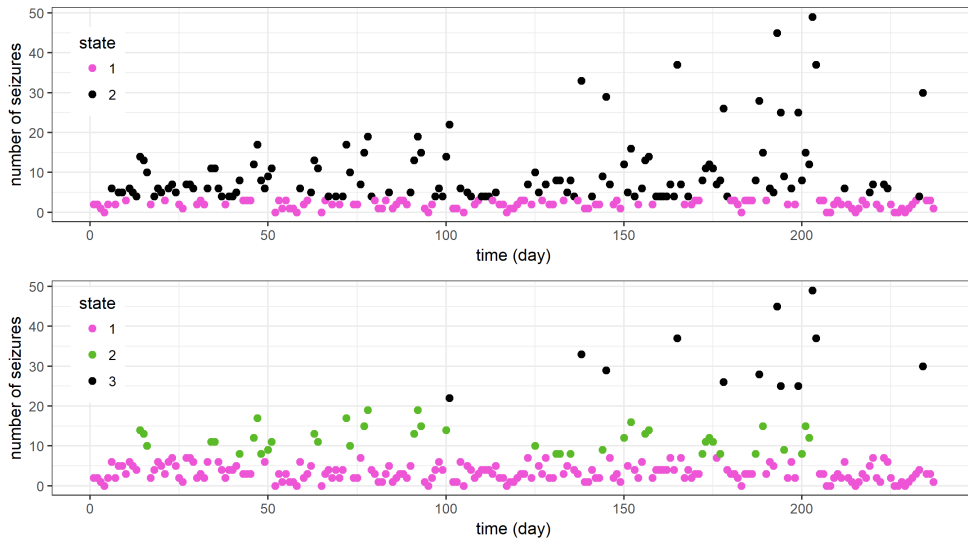


Fig. 4: Observed daily seizure counts and the estimated states assuming a mixture model with different states of epileptic activity (top: two states, bottom: three states). The mixture model assumes that the states are independent.

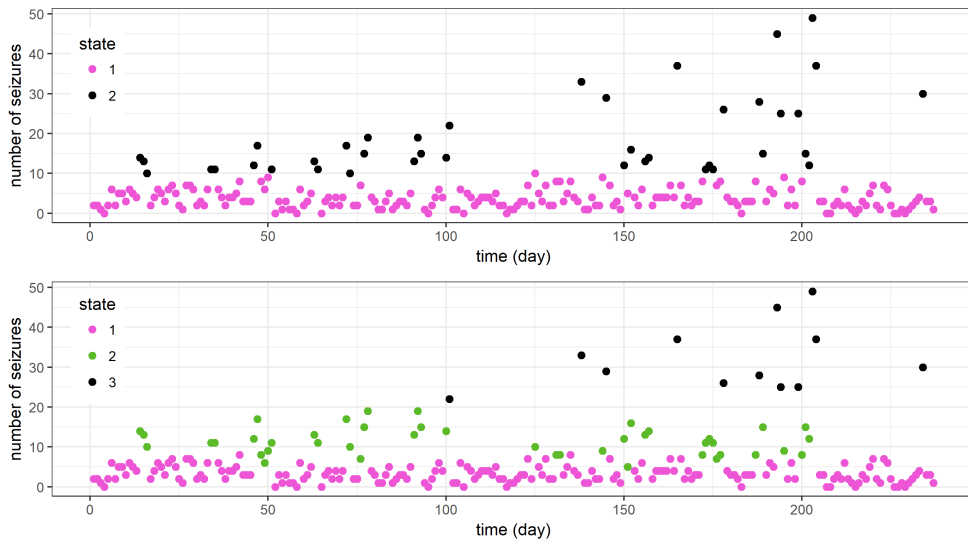


Fig. 5: Observed daily seizure counts and the estimated states assuming different states of epileptic activity (top: two states, bottom: three states). A HMM describes the dynamics of the states.

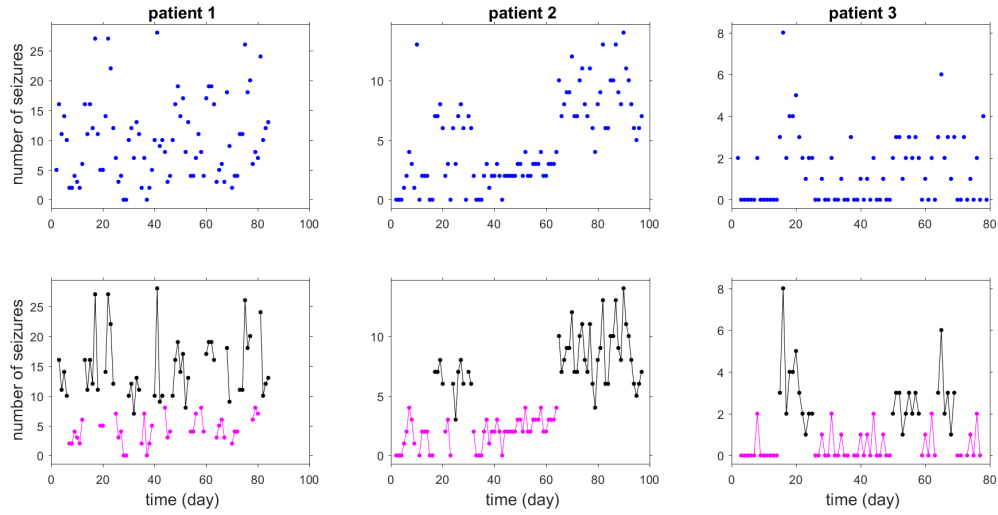


Fig. 6: Top: observed seizure counts for three typical subjects ; bottom: Estimated sequences of hidden states for the three selected subjects obtained with the Viterbi algorithm ; *black*: states of high epileptic activity, *pink*: states of low epileptic activity

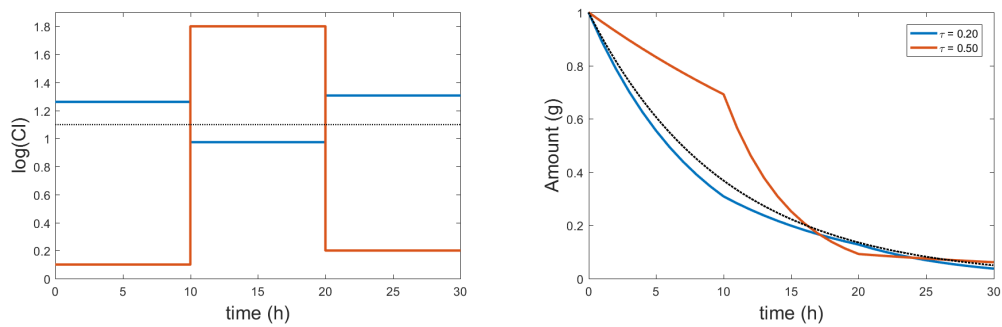


Fig. 7: Left: two sequences of log-clearances ($\log(CI_m)$, $1 \leq m \leq 3$) simulated around $\log(CI^*) = 3$ with an inter-occasion variability model and different variances τ^2 ; right: predicted amounts for these two sequences (the amount for a constant clearance $CI^* = 3$ is displayed in black dotted line).

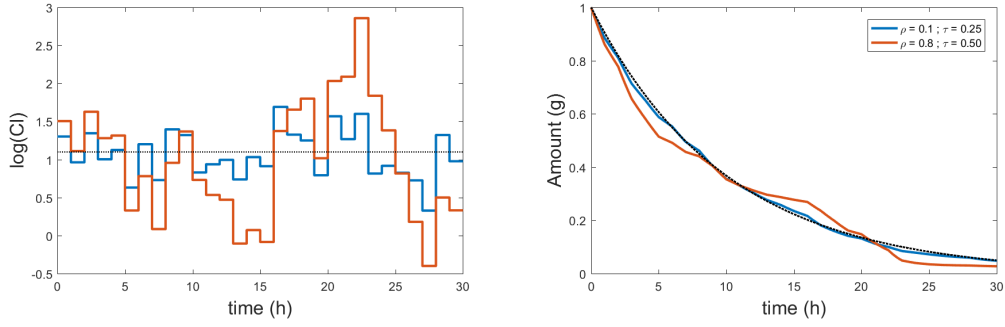


Fig. 8: Left: two sequences of log-clearances ($\log(Cl_j), 1 \leq j \leq 30$) simulated around $\log(Cl^*) = 3$ with an autoregressive model of order 1, different correlations ρ and different variances of innovation τ^2 ; right: predicted amounts for these two sequences (the amount for a constant clearance $Cl^* = 3$ is displayed in black dotted line)

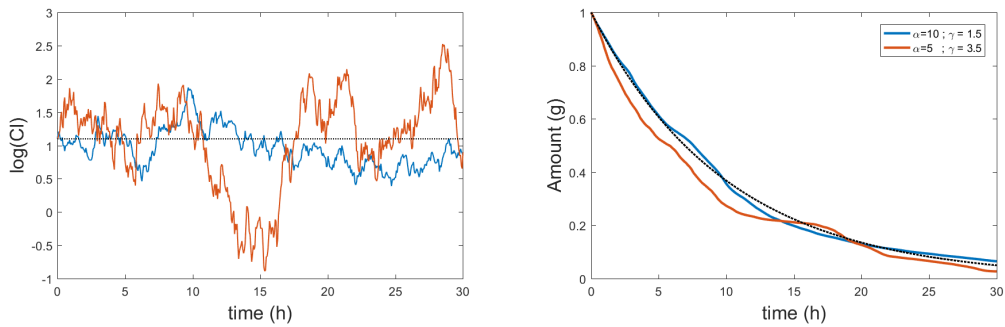


Fig. 9: Left: two sequences of log-clearances ($\log(Cl(t)), 0 \leq t \leq 30$) simulated around $\log(Cl^*) = 3$ with a Ornstein-Uhlenbeck diffusion model, different drift parameters α and different noise variances γ^2 ; right: predicted amounts for these two sequences (the amount for a constant clearance $Cl^* = 3$ is displayed in black dotted line)

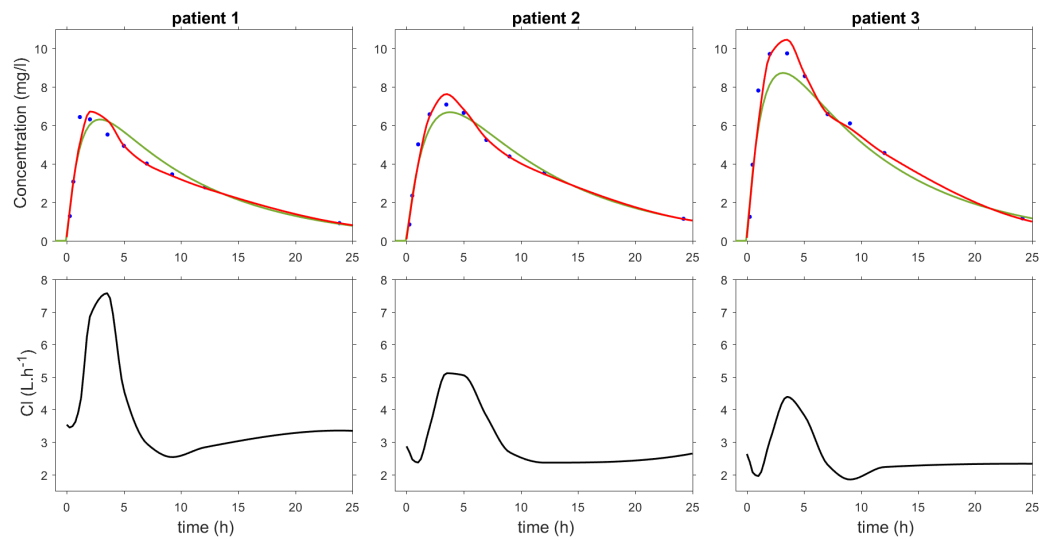


Fig. 10: Top: observed plasmatic concentrations of theophyllin (blue dots) for three subjects and their predicted concentration profiles given by the ODE based model (green) and the SDE based model (red) ; Bottom: predicted evolution of the clearance over time given by the SDE based model.

Supplementary Material

HMD_1.zip: Monolix projects (Monolix version 4.3.3) for the epileptic seizure count data example: `seizures0_project.mat` is the project file for the mixture model implemented in `hmm0_mlx.m`, `seizures1_project.mat` is the project file for the hidden Markov model implemented in `hmm1_mlx.m`. The Forward-Backward and Viterbi algorithms are implemented in `hmm1_mlx.m`. Figure 6 is produced with the m-file `seizures_plot.m`.

HMD_2.zip: Monolix projects (Monolix version 4.3.3) for the theophyllin example: `theophyllin_ODE_project.mat` is the project file for the ODE based model (1 compartment PK model with linear elimination), `theophyllin_SDE_project.mat` is the project file for the SDE based model implemented in `ekf_PKmodel.m`. The extended Kalman filter and the smoother are implemented, respectively, in `ekf.m` and `ekf_ind.m`. Figure 10 is produced with the m-file `theo_plot.m`.